Visual Reference and Iconic Content

Santiago Echeverri*†

Evidence from cognitive science supports the claim that humans and other animals see the world as divided into objects. Although this claim is widely accepted, it remains unclear whether the mechanisms of visual reference have representational content or are directly instantiated in the functional architecture. I put forward a version of the former approach that construes object files as icons for objects. This view is consistent with the evidence that motivates the architectural account, can respond to the key arguments against representational accounts, and has explanatory advantages. I draw general lessons for the philosophy of perception and the naturalization of intentionality.

In the last several decades, an impressive number of findings have accumulated to the effect that humans and other animals perceive the world as divided into objects. The evidence comes from different research traditions: developmental studies on object cognition in human infants, studies on nonhuman animals, and experiments on visual segmentation and tracking in human adults. Although there is no demonstrative argument to the effect that these researchers have been studying the same mechanisms, there are reasonable grounds to think so: this hypothesis has allowed each research community to design new experiments and offer novel solutions to account for apparently puzzling results (Leslie et al. 1998; Tremoulet et al. 2000; Carey and Xu 2001; Pylyshyn 2003, 2007; Scholl 2007; Carey 2009).

This work has been guided by the widespread consensus that object perception is underwritten by domain-specific, innate, and cognitively impen-

Received November 2015; revised July 2016.

*To contact the author, please write to: Department of Philosophy, College Avenue Campus, 106 Somerset St.—5th Floor, Rutgers University, New Brunswick, NJ 08901; e-mail: santiagoecheverri@hotmail.com.

†I presented an earlier version of this article at the 43rd Annual Philosophy of Science Conference in Dubrovnik (Croatia). I am grateful to the audience for their comments, especially to Mohan Matthen. I am also indebted to three anonymous referees for their constructive and challenging comments on earlier drafts of this article. This work was funded by the Swiss National Science Foundation (research grant 100012-150265/1).

Philosophy of Science, 84 (October 2017) pp. 761–781. 0031-8248/2017/8404-0008$10.00

etrable mechanisms. Nevertheless, there is a lively debate concerning the nature of those mechanisms. Whereas developmental psychologists characterize object segmentation and tracking as representational (Spelke 1988, 1994; Carey 2009), it has been argued that object perception is directly implemented in the functional architecture (Pylyshyn 2003, 2007). On the architectural view, visual reference is 'direct' because it is a purely causal process devoid of representational content. Hereafter, I refer to the two approaches as the representational (RA) and the architectural (AA) accounts respectively.[1]

This article defends a version of RA. The view I recommend holds that object representations have iconic content. I dub this the iconic-content account (ICA). According to Fodor (2008, 173), a content is iconic if and only if it satisfies the Parts Principle: if P is a picture of X, then parts of P are pictures of parts of X. The suggestion is that object representations satisfy this principle. They are icons that represent parts of objects arranged in some specific ways.

I doubt that one could provide a demonstrative argument to the effect that visual reference has representational content. I do think, however, that one can compare the strengths and weaknesses of representational and antirepresentational views and then decide which view is best overall. This article exemplifies this strategy. First, I submit that ICA is consistent with the key empirical evidence that motivates AA. Indeed, ICA fits that evidence even better than AA. Second, I argue that ICA can accommodate the theoretical considerations that motivate AA. Third, I contend that ICA has two explanatory advantages over AA.[2]

Although the main focus will be on the philosophy of cognitive science, the discussion will enable us to draw lessons of broader significance to philosophy. These lessons touch on relational theories of perceptual experience, the characterization of the structure of perceptual content, and the project of naturalizing intentionality.[3]

The article proceeds as follows. After introducing some conceptual distinctions (sec. 1), I present the two leading accounts of visual reference in recent cognitive science: RA and AA (sec. 2). Next, I sketch ICA (sec. 3). I defend ICA in the remainder of the article (secs. 4–7).

1. Conceptual Distinctions. A number of researchers have introduced the concept of a mental file to frame their theories of object cognition and refer-

1. I provide a precise characterization of these views in sec. 2.

2. Carey (2009, 68, 135ff.) has also formulated the conjecture that object representations have iconic content. Nevertheless, she has not motivated this view in her analysis of visual reference.

3. Echeverri (2016a, 2016b) explores other philosophical ramifications of the current approach.

ence (Kahneman, Treisman, and Gibbs 1992; Dickie 2010; Recanati 2012). A mental file is a mental representation that enables a system to store information about individuals. I will adopt this framework here.

I am interested in the ability to visually refer to objects like trees, tables, and rocks. Most cognitive scientists hold that visual reference is underwritten by two largely different abilities: segmentation and recognition (Kahneman et al. 1992, 176–77). *Segmentation* is the ability to 'detach' various elements from the scene as one object. This ability differs from *recognition* because one can segment unfamiliar objects. Think of an astronaut exploring the surface of an unknown planet. Although she might be unable to recognize any object, she could certainly segment some objects from the rest of the scene. I will focus here on the ability to segment visually.

We can also use the segmentation/recognition contrast to distinguish two kinds of mental files: object files and recognitional files (Kahneman et al. 1992, 176–77). Suppose that our astronaut found a very strange object. When this occurred, an object file was activated in working memory to represent the strange object. If she managed to recognize that object, a recognitional file containing long-term memories would become active. This file might include information about that specific object or about the sort of object it is. Since I am interested only in the mechanisms that underlie the activation and maintenance of object files, I remain neutral on the nature of recognitional files.

It would take us too far afield to provide an analysis of the concept of an object. To circumvent this problem, I follow the standard practice of characterizing objects by means of examples. Paradigmatic objects include trees, tables, and rocks. Many cognitive scientists think that our ability to segment objects is partly determined by a number of principles (Spelke 1994; Carey 2009). I call these principles 'object constraints' because they impose conditions that some aggregates of elements must satisfy in order to be parsed as objects. Here are three constraints:

> *Continuity*: Humans (and other animals) attend to and keep track of entities that seem to follow a continuous path through space and time.
>
> *Cohesion*: Humans (and other animals) understand objects as entities that keep their parts bound together, forming an internally unified whole.
>
> *Permanence*: Humans (and other animals) understand objects as entities that continue to exist even if they are out of sight. (Spelke 1994; Carey 2009)

Although objects such as trees, tables, and rocks respect these constraints, satisfying them is insufficient for an entity to count as a *material* object. After all, entities we do not usually count as material objects—such as circles

on screens—may behave in ways that respect these conditions. Let us introduce the expression 'visual object' to denote aggregates of entities that satisfy the object constraints but are not paradigmatic material objects. A circle on a screen is different from paradigmatic material objects because it is not three-dimensional and has a very ephemeral existence. Still, a circle on a screen can respect some object constraints in specially contrived experimental setups (Pylyshyn 2003, 2007; Matthen 2005; Scholl 2007). Thus, a circle on a screen can lead to the formation of an object file. Hereafter, when I speak of 'objects', I mean either material or visual objects.

I remain neutral on the relations between material and visual objects. Nevertheless, I impose a condition on the referents of object files: they are mind-independent entities. There are three main motivations for this condition. First, if object files can misrepresent their objects, the latter should be characterized independently of our visual or cognitive abilities. Thus, this assumption is essential to any representational account of visual reference. Second, this approach fits pretty well with the widespread belief that our abilities to segment and track objects have increased the fitness and reproductive success of our ancestors (Millikan 2000; Pylyshyn 2003, 2007; Carey 2009). Intuitively, it is the mind-independent moving target that is tracked by the lion and serves it as a basis for inductive learning. Third, the design of the experiments requires a characterization of the targets of object files that is independent of subjects' specific cognitive and perceptual abilities. If this characterization were not available, it would be impossible to determine whether they successfully completed the task.[4]

2. Two Accounts. A number of developmental psychologists have characterized object perception as underwritten by representational systems (Spelke 1988, 1994; Carey 2009). There are at least two ways of fleshing out this claim. First, one might hold that visual reference involves the operation of object representations such as object files (sec. 1). Indeed, Carey (2009) and her colleagues (Carey and Xu 2001) have made a strong case for the claim that object files serve as input for physical reasoning, enter into number-relevant computations, and support intentional action. There is, however, a more controversial way of fleshing out this approach:

> *Representational Account* (RA): The mechanisms that underlie the activation and maintenance of object files have representational content.

This view is explicit in Carey's (2009, 68, 135ff.) conjecture that object representations have iconic content. It is also implicit in Spelke's (1988) conjecture that object cognition is underwritten by a conceptual system.

4. I come back to this point in sec. 4.

Pylyshyn (2001, 2003, 2007) has rejected RA. Although he has used the mental-file framework in order to explain visual reference, he also thinks that the mechanisms that activate and maintain object files refer in a causal, nonrepresentational manner: "This sort of causal connection between a perceptual system and an object in a scene is quite different from a representational or intentional or conceptual connection. For one thing there can be no question of the object being *mis*represented since it is not represented *as* something" (Pylyshyn 2001, 147; see also his 2003, 219 n. 7; 2007, 49).

Contrary to developmental psychologists, Pylyshyn thinks that the mechanisms underlying the activation and maintenance of object files are directly implemented in the functional architecture (AA). He defends this approach by denying RA:

> *Nonrepresentational Account*: The mechanisms that underlie the activation and maintenance of object files lack representational content.

Pylyshyn's reasoning requires some unpacking. In computer science, the concept of a functional architecture characterizes the *elementary computational resources* that are necessary to realize cognitive processes (Pylyshyn 1984, 30, 154, 161; 2001, 148). Thus, if one holds that segmentation and tracking are part of the functional architecture, one is claiming that segmentation and tracking are elementary computational resources (Pylyshyn 2003, 258; 2007, 39).

Pylyshyn (1984) is committed to a semantic view of computation. He thinks that computations are defined over mental representations (see Piccinini [2015] for an overview). Thus, if an operation is directly instantiated in the functional architecture, it is not computational, at least on a semantic understanding of computation. It is "simply performed, or 'instantiated', by properties of the biological substrate in a manner not requiring the postulation of internal representations. To ask how a primitive function is carried out is to ask for a functional or perhaps a physical (or biological) description" (Pylyshyn 1984, 154; see also xvii, 30–31, 132–33).

Suppose one holds—as Pylyshyn does—that the aim of cognitive science is to explain complex abilities by formulating algorithms that operate over mental representations. On this view, any phenomenon that cannot be explained via an algorithm that operates over mental representations will automatically fall outside the purview of cognitive science. This is certainly a controversial view. Still, it reveals what is at stake in his debate with developmental psychologists. If being representational is sufficient to fall within the purview of cognitive science, developmental psychologists have been trying to elucidate visual reference within cognitive science. Pylyshyn insists, by contrast, that it is not the business of cognitive science to explain visual reference.

Pylyshyn's defense of AA is based on an inference to the best explanation. He thinks that AA best explains the available empirical evidence and some theoretical requirements on the mechanisms that activate and maintain object files. I will dispute these claims. Indeed, there is at least one version of RA that is consistent with the empirical evidence, can respond to Pylyshyn's theoretical considerations, and has some explanatory advantages over AA. Before I present these arguments, I have to introduce the version of RA I recommend.

3. The Iconic-Content Account. Pylyshyn's central claim is that the reference of object files "is quite different from a representational or intentional or conceptual connection." The version of RA I recommend holds that the connection of object files with objects is representational albeit not conceptual. What does it mean to hold that visual reference is representational?

Let us distinguish a representation from its representational content. The string 'Pierre can fly'—phonologically or morphologically individuated—is a representation. By analogy, its psychological counterpart PIERRE CAN FLY is a *mental* representation. Let us say that a mental representation carries a representational content. A representational content is an abstract entity. Typical examples of contents are propositions and their constituents. We will say that representational contents determine *correctness conditions*. The latter are situations under which a representational content is correct or incorrect. The proposition ⟨Pierre, ⟨can fly⟩⟩ is correct if and only if Pierre can fly. It is incorrect otherwise. Let us use the adjective 'correct' as a generic term covering a variety of evaluations like truth or accuracy. The claim that Pylyshyn rejects but I defend is that abstract entities that determine correctness conditions mediate the reference of object files.

To be sure, Pylyshyn does admit that object files can be constituents of complex representations that carry representational contents. He dubs the mechanisms that fix the reference of object files 'FINSTS'. These devices work in a similar way to linguistic demonstratives, for they can refer to different objects in different perceptual contexts (Pylyshyn 2007, 16 n. 5). Thus, when the visual system encodes the property blue, Pylyshyn will interpret this as the tokening of the representation THAT IS BLUE, a representation that can be assessed as correct or incorrect. What Pylyshyn denies but I accept is that the object file itself—before any perceptual attribution to an object—has representational content and that this content mediates its reference.

This proposal may seem counterintuitive. If we follow the linguistic analogy too closely, tokens of 'that' cannot have representational content unless they are paired with a predicate, for only their relation to a predicate will enable them to represent its referent as something (Pylyshyn 2003, 219 n. 7). Nevertheless, my suggestion will gain plausibility if we think of object files as icons that carry iconic content. A picture can accurately represent your

father even though it lacks predicative structure. Thus, object files can be correct or incorrect in the same way in which pictures can be correct or incorrect. According to Fodor (2008, 173), a content is iconic if and only if it satisfies the Parts Principle: if P is a picture of X, then parts of P are pictures of parts of X. The idea is that object files are icons analogous to pictures and different from words. A picture of a dog represents not only the dog but also its ears, its eyes, its nose, and other dog parts. In contrast, the word 'dog' does not represent any part of the dog (Carey 2009, 135). This is the core idea of ICA: object files are icons that represent objects and their parts (97ff.).[5]

ICA has a straightforward consequence for the project of explaining visual reference. Instead of asking questions concerning the 'activation' of the quasi-linguistic symbol THAT, we should rather spell out the conditions under which the visual system builds up icons for objects. Of course, this is not to deny that object files may be similar to demonstratives in some respects. The point is that the linguistic analogy masks the representational complexity of object files.

We can now distinguish AA from ICA. AA holds that visual reference is achieved via a semantically primitive demonstrative that directly picks out objects. ICA grants that we may need to posit semantically primitive representations. Still, these semantically primitive representations refer not to objects but to the pre-objective elements used by the visual system as input to build up object files. By a 'pre-objective element' I mean any property instance that can be used as input to build up object representations. They will include slant, vertical, and horizontal bars, vertexes, T-junctions, chromatic properties, and so on. Hereafter, I use 'thin index' to denote any device that refers to a pre-objective element.

Thin indexes, however numerous, are not sufficient to produce a representation of an object. They would be sufficient only if objects were bare collections of elements. Objects are not bare collections of elements but highly structured entities. That is why one and the same collection of elements can form different objects. For this reason, if the visual system can form icons for objects, it will need some *combinatorial principles* that confer structure on active thin indexes. These principles will enable the visual system to distinguish different objects formed by the same collections of elements. ICA hypothesizes that so-called object constraints are combinatorial principles that govern the formation and maintenance of object files.

5. We should be careful to distinguish Fodor's Parts Principle from another, false principle: if $E_1, E_2, \ldots, E_{n-1}, E_n$ are parts of X, and a system has a picture P of X, then there are parts of P that stand for $E_1, E_2, \ldots, E_{n-1}, E_n$. This principle precludes the possibility of having sparse iconic representations, i.e., iconic representations that represent some but not all of the proper parts of an object. See Matthen (2014, 277).

ICA says that the mechanisms of object segmentation and tracking manipulate thin indexes in conformity with object constraints.[6]

We have now the materials to characterize the correctness conditions of object files. Object files are governed by a 'matching condition'.

> *Matching Condition*: If a sequence of active thin indexes $I_1, I_2, \ldots, I_{n-1}, I_n$ are combined in conformity with an object constraint, *C*, then there is one combination of pre-objective elements $E_1, E_2, \ldots, E_{n-1}, E_n$ that stand in an appropriate causal relation to $I_1, I_2, \ldots, I_{n-1}, I_n$ and exemplify the object constraint *C*.

It has been observed that perceptual content has accuracy conditions because it can be correct to some degree (Burge 2010). We could therefore refine the matching condition in order to make room for partially accurate iconic contents. One might count the visual system as successfully segmenting an object even though not all the thin indexes over which object constraints operate select elements belonging to one and the same object. Think of the visual segmentation of objects in cluttered scenes. If objects are in contact with each other, the visual system may be led to form one icon that includes some proper parts of numerically different neighboring objects. Still, within some limits, this might not impair successful segmentation. The degrees of inaccuracy that are still compatible with successful segmentation may depend on the task at hand and the perceptual context. Their determination is an empirical matter.[7]

Recall now Pylyshyn's contention that the "causal connection between a perceptual system and an object in a scene is quite different from a representational or intentional . . . connection." ICA rejects this claim because it can define an iconic content for object files. Consider the following extensional characterization. The iconic content of an object file $|O_x|$ is a function $f(|E_1, E_2, \ldots, E_{n-1}, E_n|)$ of a sequence $|E_1, E_2, \ldots, E_{n-1}, E_n|$ of pre-objective elements. When the argument of the function is given, the icon O_x is formed. Yet, the formation of O_x is not sufficient for successful reference. An active icon O_x successfully refers to an object in the world only if there is one sequence $|E_1, E_2, \ldots, E_{n-1}, E_n|$ of pre-objective elements $E_1, E_2, \ldots, E_{n-1}, E_n$ in the world that matches and stands in an appropriate causal relation to the represented sequence. (Of course, we could refine this characterization in or-

6. Object constraints should be supplemented with other organizational principles, such as those governing the figure-ground distinction. These principles enable us to perceive some surfaces in depth, so they are necessary to perceive material objects. See Echeverri (2016b).

7. I come back to this point in sec. 4.

der to make room for cases of partial matching that are still compatible with successful reference).[8]

Pylyshyn accepts that an object file can carry representational content, provided that it is attached to a predicate. The representation THAT IS BLUE can be correct or incorrect. Yet, Pylyshyn denies that a bare occurrence of THAT can have representational content. ICA resists this conclusion. Indeed, it recognizes a whole class of 'pre-attributional' perceptual errors. Consider an example.

> *Binding proper parts of numerically different objects into one object*. Imagine that you take two lemons of exactly the same size, texture, and color. You cut each lemon into two halves. Next, you put one-half of lemon 1 with another half of lemon 2 so that the flesh of the former makes contact with the flesh of the latter. You might then put the two halves on a table and eventually 'fool' someone into seeing the two halves as one object. This case would count as a misrepresentation. After all, the two halves are not disposed to form an internally unified whole, contradicting the principle of cohesion (sec. 1).

ICA provides an elegant characterization of this case. Although the perceptual episode may involve the misattribution of some properties to the whole of two lemon halves, there is a kind of error that does not consist in the misattribution of properties to that entity. After all, in order to misattribute any property to any object, one has first to successfully segment it. Thus, the previous case illustrates the idea that object files, understood as icons, can misrepresent objects prior to any attribution of properties to objects. If one follows Pylyshyn's lead, however, one will be compelled to treat all segmentation mistakes—if there are any—as attribution mistakes. Although I see this as a disadvantage of Pylyshyn's view, it will require some work to show why it is. I do that in the reminder of the article.[9]

8. This extensional specification is not mandatory. We might also specify the iconic content via nonextensional entities analogous to Fregean senses that stand for pre-objective elements. This maneuver would enable us to assign a common content to perception and perceptual imagination.

9. Preattributional mistakes are more pervasive than the lemon example might suggest. Nevertheless, their existence may be masked because they tend to co-occur with attributional errors downstream of visual segmentation. Suppose that you are walking in a park and seem to see a small dog sniffing something under a distant bush. On closer approach you see that it was just part of a paper bag, some branches, and a play of shadow. In this case, a preattributional mistake led you to misrecognize a portion of reality as a dog. It is also plausible to hold that impossible figures are preceded by segmentation mistakes. The reason why the visual system ends up attributing incompatible properties to the same thing is that it made a segmentation mistake. The empirical literature on visual agnosia might also provide indirect evidence of preattributional perceptual errors, for patients suffering this deficit have impaired segmentation abilities (see Humphreys and Riddoch 2014). I am indebted here to a referee for this journal.

4. Empirical Evidence. My first argument is that ICA successfully accounts for the empirical evidence that originally motivated AA. Indeed, it offers a superior account to the one provided by AA.

Pylyshyn and his coworkers created an experimental paradigm in order to test AA: multiple-object tracking (MOT). In a typical MOT experiment, subjects are presented with various qualitatively identical items on a computer screen. Some of them are flashed to indicate their status as targets. After that, the objects start moving randomly for some time (usually 10 seconds). At the end, observers are asked to indicate the original targets. These studies suggest that humans can keep track of four or five moving targets in parallel. Some variations of this paradigm also suggest that observers can track individuals even though they change some of their properties, like shape and color. Strikingly, subjects are often unaware of changes in objects' properties during tracking. As Pylyshyn puts it: "It appears that nothing is stored in the object files under typical MOT conditions, which suggests that targets are not being picked out under a description—they are not picked out as things that have certain properties or satisfy certain predicates" (2007, 40; see also 21, 34 52, 68).

In order to track an object, the visual system traces its numerical identity "back to a prior state when it was known to be a target" (Pylyshyn 2007, 45). Hence, visual tracking is governed by a principle of continuity. For our current purposes, the most striking aspect of Pylyshyn's findings is that representations of the shapes and colors of objects do not seem to play any role in the maintenance of object files.[10]

Pylyshyn appeals to work in the philosophy of language to interpret these results. Philosophers have explored two broadly different accounts of how singular expressions refer. According to descriptivism, reference is fixed by an object's satisfaction of properties. According to referentialism, reference is fixed by causal relations to objects (see Recanati [2012] for discussion). Given that representations of the color and shape of objects do not seem to play a role in tracking, Pylyshyn (2007, 17–19, 68) concludes that a descriptivist account of visual tracking is not an option. Assuming that descriptivism and referentialism exhaust the theoretical options, he is led to hypothesize that the latter offers the best framework to account for these empirical findings. On this view, the reference of object files is maintained by a causal, nonrepresentational mechanism.

Pylyshyn's contention that representations of the properties of objects play no role in successful tracking is also based on the idea that targets and nontargets are qualitatively indistinguishable from each other. One might think that this offers an argument to reject ICA, for the latter predicts that the en-

10. I qualify this interpretation below. Pylyshyn also insists that representations of locations and sortals play no role in visual tracking. I will not discuss these claims here.

coding of some property instances is necessary for successful tracking. After all, if object constraints are combinatorial principles, they need active thin indexes as inputs. And these thin indexes refer to property instances.

This conclusion is, however, mistaken. If targets lacked some detectable properties, they would blend into the background; they would not be salient as targets. Therefore, even if properties are not necessary to distinguish each target from other targets and nontargets, they are still necessary to distinguish each target from the background. ICA accommodates this insight by saying that thin indexes that refer to property instances must remain active during visual tracking. Yet, it is irrelevant to successful tracking whether this thin index or that thin index becomes active. All the visual system needs is that some thin indexes detect some property instances. Successful tracking depends on whether these thin-indexed pre-objective elements display the right patterns of behavior over time (i.e., the sorts of behavior described by object constraints).

Consider an analogy. Suppose that you are watching a group of people jogging very far away on a field. In order to keep track of the group, each member should have at least some detectable chromatic properties that distinguish it from the field. Yet, your ability to keep track of the group does not require the additional ability to keep track of the chromatic properties of each jogger. Given the distance, you may be unable to determine whether member 1 has the same chromatic properties as member 2 or whether member 1 changed her chromatic properties from t_1 to t_2. Still, this will not prevent you from successfully tracking the group. All you need is that each jogger has at least some property instance that your visual system can detect, enabling you to distinguish it from the field. What enables you to keep track of the group is your ability to detect a sustained pattern of behavior defined over chromatically discriminable members of the group, independently of which chromatic properties are enabling you to make the target/background differentiation. The proposal is that object files work in a similar way. They are icons that collect pre-objective elements that satisfy some constraints. Because these collections satisfy object constraints, they differ from groups of people, which display a lower degree of cohesiveness and continuity over time.

Why do subjects fail to notice some changes of properties? According to an influential view, working memory is necessary for access (Prinz 2012). Recall now that object files are standardly construed as representations in working memory (sec. 1). Thus, one might submit that the property instances used as input to build up object files are not encoded in working memory. In Pylyshyn's framework, in order for a property F to be encoded in working memory, the tokening of a representation of the form THAT IS F is necessary. Interestingly, ICA offers a representational structure that precedes the formation of these predicative representations. Property instances only figure as proper parts of represented objects, not as the semantic values of predicates of object

files. Therefore, ICA can explain why subjects fail to report some changes of properties.

Interestingly, Pylyshyn is aware that some results of MOT experiments run against his indexing theory. I want to suggest now that ICA is better placed than AA to explain these results.

It has been observed that increasing the number of nontargets will impair performance (Sears and Pylyshyn 2000; Pylyshyn 2004). This contradicts the FINST theory, which says that successful tracking is solely determined by a causal relation that holds between FINSTS and individual targets. For this reason, Sears and Pylyshyn (2000) have modified the FINST theory accordingly: increasing the nontargets in a given display will raise the probability of mistaking a nontarget for a target. As Pylyshyn puts it: "Confusing one individual object with another object represents a *failure to correctly track* that object. . . . Switching the identity of a target with that of a nontarget does show up as *a tracking error*" (2004, 816; emphasis mine).

Pylyshyn seems to be moving here toward a RA of visual tracking. Interestingly, whereas ICA does predict that these tracking errors should occur, AA does not predict their occurrence. Therefore, ICA is better placed than AA to account for these findings.

Recall that ICA imposes a matching condition on successful visual reference: if the thin-indexed pre-objective elements do not belong to one combination of elements in the world, the iconic representation will be (partially) inaccurate (sec. 3). Thus, when the number of nontargets is increased, this raises the probability that some of the thin indexes that are attached to a target become attached also to a nontarget. If all the thin indexes initially attached to a target switch to a nontarget, the visual system will be led to misrepresent a nontarget as numerically the same as the target or the nontarget as a temporal continuation of the target. This is an example of preattributional misrepresentation. One object file O_x remains active during the whole experiment. Still, some of the elements E_1, E_2, E_3 that work as inputs for that object file belong to the target, while other elements E_4, E_5, E_6 belong to a nontarget. Assuming that all these elements are linked to each other by the principle of continuity, this is a case of misrepresentation. The visual system is led to 'merge' into one sequence what are in fact two different sequences. It merged parts of the 'target worm' and parts of the 'nontarget worm' into a single spatiotemporal worm.

Some might wonder whether Pylyshyn could not account for these results without introducing object-file misrepresentation.[11] I am skeptic about the prospects of this view.

11. I am indebted here to Mohan Matthen. As far as I can see, this constitutes the most natural line of reply available to Pylyshyn. There could be other replies, though. I leave them as an exercise for the reader.

Pylyshyn should grant that the task at hand sets some correctness conditions. The MOT task is successfully performed just in case the items that were indicated as targets at the beginning of the experiment match the items that subjects indicate as targets at the end of the experiment. Thus, even defenders of AA should grant that tracking is evaluable as correct or incorrect. Still, they could try to dissociate the assessment of the task as correctly or incorrectly performed from the introduction of incorrect mental representations.

In my view, this move would yield a complicated and ad hoc account. First, Pylyshyn would have to grant that the indexes that were causally covarying with a target started to covary with a nontarget. Second, Pylyshyn would have to explain why this covariation shift occurred. One option would be to offer no account of the conditions under which targets coming close to nontargets swap indexes. This would represent a disadvantage of AA over ICA. After all, ICA says that those switches arise when elements of targets and nontargets satisfy—for a short time at least—some object constraints. In the current case, pre-objective elements belonging to different objects are seen as a continuous spatiotemporal worm. Another option would be to classify these cases into two groups: those in which a target coming close to a nontarget leads to a swap of indexes and those in which a target coming close to a nontarget does not lead to a swap of indexes. In this case, we would like to know what principles enable defenders of AA to distinguish these two kinds of cases. If the principles are nothing but object constraints, Pylyshyn's account is a notational variant of ICA. If they are not object constraints, the resulting account is more complicated than ICA. AA is led to posit—in an ad hoc manner—object constraints to account for successful tracking and other principles to account for target/nontarget switches.

5. The Regress Argument. My second argument is that ICA can respond to the regress argument, the key theoretical consideration in favor of AA. Pylyshyn (2003, 245) criticizes what he terms the 'pure description' theory of reference. His critique consists of two claims:

1. Pure description theories of reference lead to a regress.
2. The best way of blocking the regress is to hold that the visual system can select objects in a purely causal manner.

According to Pylyshyn's (2003, 205, 245; 2007, 12ff.) characterization, pure descriptivism explains visual reference by positing a set of predicates that represent properties. Let us suppose that an object file is associated with some predicates $F_1(x)$, $F_2(x)$, . . . , $F_{n-1}(x)$, $F_n(x)$. The relevant object file refers to whatever entity satisfies $F_1(x)$, $F_2(x)$, . . . , $F_{n-1}(x)$, $F_n(x)$, most of them, or a weighted sum of them. As Pylyshyn points out, this view leads to a re-

gress. Suppose we try to explain how the predicate $F_i(x)$ is anchored to an object o. In order to yield a representation of the form $F_i(o)$, the process could not rely on a further predicate $F_{i+1}(x)$. After all, if the selection of an object required a prior application of the predicate $F_{i+1}(x)$, we would get $F_i(F_{i+1}(x))$. If the only way of referring to an object were to encode a property, we would be launched in a regress. Since there is no such regress—because the visual system does refer to objects—pure description theories are false. There must be nondescriptive modes of reference.

What is the nature of these nondescriptive modes of reference? As indicated in section 4, Pylyshyn appeals to work in the philosophy of language in order to develop his own views. Following the lead of philosophers of language, Pylyshyn concludes that nondescriptive modes of reference should be modeled along referentialist lines (i.e., as causal relations to objects). Pylyshyn's suggestion is that there is a causal relation that fixes the reference of object files before any representation of properties. Standing in a causal relation to an object is a precondition for representing the properties of that object (Pylyshyn 2003, 2007; see also Leslie et al. 1998; Campbell 2012; Recanati 2012).

But why should we think of that mechanism as a causal relation to objects? Pylyshyn motivates this view with a different regress argument: "Sooner or later concepts must be grounded in a primitive causal connection between thoughts and things. The project of grounding concepts . . . in perception remains an essential requirement if we are to avoid an infinite regress" (2001, 154; see also his 2007, 17, 33, 57).

Pylyshyn supports his referentialist view by combining two different regress arguments. The first one establishes the existence of nondescriptive modes of reference to objects. The second one establishes that there cannot be mental representations all the way down; any account of reference has to introduce primitive causal connections with items in the world. Thus, in Pylyshyn's framework, primitive causal connections are introduced at the level of nondescriptive modes of reference to objects.

ICA resists this identification of the regress arguments. Although it grants that any theory of mental representation must be grounded in primitive causal connections with items in the world, it denies that those primitive causal connections ought to hold between thoughts and things or, more precisely, between object files and objects. Indeed, if those connections hold between some thin indexes and pre-objective elements in the world, one can hold that nondescriptive modes of reference have representational content without being prey to the second regress argument.[12]

12. The qualification 'some thin indexes' is important because the visual system may confer a rudimentary structure to representations of clusters of features. Rensnik's (2000) hypothesis of 'proto-objects'—volatile clusters of features that fade away within some milli-

The preceding discussion has wider implications. Philosophers of perception have used regress arguments structurally analogous to the ones examined here in order to defend the claim that nonrepresentational relations to objects play a fundamental role in the analysis of the structure of perceptual experiences (e.g., Campbell 2009). If we make room for the existence of iconic representational contents, we can see that there is a non sequitur in these arguments.

6. The Grounding Problem. Pylyshyn also thinks that it is unclear how RA could ground object representations. My third argument is that ICA can offer a plausible account of how object representations are grounded.

Commenting on RA, Pylyshyn writes:

> The question of whether infants have the concept of object . . . does eventually run into the need to ground that concept in experience (by 'ground' I mean connect the concept with its instances, not *learn* or acquire the concept, which may well be innate). For example, it has been suggested that the first sortal concept that an infant has is the concept *object* . . . , which is the concept of something that is 'bounded, coherent, three-dimensional, and moves as a whole' . . . . But of course if that's what an object is for an infant, then infants must also have the concepts 'bounded', 'coherent', 'three-dimensional', and 'moves as a whole', in which case the Spelke object could not be their first concept. (Pylyshyn 2007, 51)

In order to interpret this paragraph, we need a distinction between two kinds of projects: the naturalization project and the grounding project. The naturalization project seeks to identify a set of naturalistic conditions that are sufficient for something to count as a representation. These conditions may invoke the concepts of information and biological function. The grounding project seeks to identify the general structure of the mechanisms that connect mental representations with their referents. Pylyshyn's remarks belong to this second project. His worry is that, if we construe object files as complex representations involving representations of the properties bounded, coherent, moves as a whole, and so on, it will be a mystery how these representations are connected with their referents. This feeling of mystery is exacerbated by the first regress argument against pure description theories (sec. 5). And, even if we set aside these views, it remains unclear how such general representations can be connected with all and only things that satisfy object constraints. If we wanted to design a mechanism that would perform this task, we would be at a loss.

---

seconds or are overwritten by subsequent stimuli—constitutes an example of this idea. Since they are not objects in the sense of sec. 1, they would count as pre-objective.

Pylyshyn's worry will have a hold on us only if we construe the properties of being bounded, coherent, or moves as a whole as represented by predicates of object files. But this assumption is not mandatory. Suppose that object files are icons. Although icons have a complex representational structure, they lack the quasi-linguistic predicative structure that is taken for granted in Pylyshyn's objection. Thus, one might hold that object files are aggregates of thin indexes organized in conformity with object constraints. Thus, the problem of how very general representations of object constraints are connected with their bearer does not arise.

In order to see why the problem does not arise, we have to improve our understanding of the available theoretical options. Let us stipulate that a property *F* is *explicitly represented* by a system, *S*, just in case *S* has a physical structure that stands for *F* and this physical structure can be used as input for other operations. Thus, some neurons explicitly represent features because they have the function of covarying with those features and can be used as input for other operations. Let us stipulate that a property *G* is *implicitly represented* by a system, *S*, just in case *G* characterizes some mental operations performed over explicit representations. Hence, a system that forms an object file in response to a sequence $|E_1, E_2, \ldots, E_{n-1}, E_n|$ is a system that explicitly represents $E_1, E_2, \ldots, E_{n-1}, E_n$ and implicitly represents the object constraints, for the latter characterize the way it combines those elements. Therefore, it is true that the satisfaction of object constraints is necessary for the formation and maintenance of object files. Nevertheless, this satisfactional relation is not psychologically implemented as a relation between an object file and a set of predicates. It is rather implemented in the dispositions of the visual system to combine pre-objective elements in some specific ways.

The resulting view preserves part of the spirit of AA. There are indexes that refer in a purely causal manner. Yet, contra Pylyshyn, those indexes directly refer not to objects but to pre-objective elements. In addition, there are operations that govern visual segmentation and tracking. Nevertheless, while Pylyshyn treats visual segmentation and tracking as basic operations of the functional architecture, ICA factorizes object segmentation and tracking as the interplay of mechanisms that refer to pre-objective elements plus a battery of object constraints that govern the dispositions of the system to combine thin indexes in some ways.

We are now in a position to see how ICA can provide a solution to the grounding problem. When we explain a piece of behavior by means of dispositions, we are committed to providing a noncircular account of their triggering conditions. If we said, "the dispositions to form an object file are triggered by the detection of an object," our account would be circular. Thus, all we need is a noncircular account of the triggering conditions of the dispositions that govern the formation and maintenance of object files. This account

can be provided. If an aggregate of elements is an object, it is likely that it instantiates a number of local configurations that are not typically instantiated by nonobjects. Therefore, ICA can solve the grounding problem by positing dispositions to combine thin indexes and hypothesizing that these dispositions are triggered by local configurations detectable by the visual system.

The edges formed by objects when they are momentarily occluded provide a telling example of such local configurations. In some experiments, observers are asked to keep track of objects in spaces filled with virtual occluders (Scholl and Pylyshyn 1999). These experiments are based on the plausible idea that objects do not pop into or out of existence, although they do frequently pop into and out of sight. Two scenarios have been tested. In some cases, the targets disappear by gradually 'imploding' and later 'exploding'. In others, they disappear in a way that indicates the presence of occluding surfaces, by accreting and deleting along a fixed contour. It has been found that tracking abilities are not impaired in the latter case.

We can generalize from these findings as follows. The edges formed by a momentarily occluded object are some of the triggering conditions of the dispositions that govern object tracking. Being sensitive to these edges is not equivalent to being sensitive to objects. Hence, if tracking is governed by dispositions that are triggered by the edges formed by momentarily occluded objects, the current account is not circular. The idea is that the visual system is sensitive to this and other local configurations that trigger object representations. Other plausible examples can be found in work on perceptual organization (e.g., Hoffman 1998).

The previous discussion has some implications for recent debates on perceptual content. According to Burge (2009, 2010), perceptual content has a structure analogous to a noun phrase of the form 'that F': 'that' signals the fact that successful perception requires a causal relation to an entity, and 'F' denotes an attribute whose representation guides reference to that entity. Thus, Burge (2010, 456 n. 39) interprets the empirical findings on object perception as evidence for the claim that visual reference is achieved via a complex demonstrative in which BOUNDED, COHERENT, or CONTINUOUS figure as attributive representations that guide the reference of THAT. Unfortunately, Burge's view seems to generate the same worries raised by Pylyshyn in his discussion of RA. How are these attributive representations connected with objects?

As far as I can see, Burge's analysis cannot jointly satisfy two plausible conditions: (1) accommodate the intuition that any constituent of a complex representation, *R*, should make a nonredundant contribution to the content of *R* and (2) explain why the associated attributives in the object representation apply to all and only things that are bounded, coherent, and so on. Burge could try to satisfy condition 1 by characterizing THAT as a placeholder for a

causal relation between the visual system and any mind-independent entity in the world. This reading is strongly suggested by the generality of Burge's analysis: noun phrase structures of the form THAT F are supposed to describe the semantic structure of perceptual representations of bodies, shapes, colors, and any other entity that can remain the same over changes in information registration. Unfortunately, this approach would prevent Burge from satisfying condition 2. After all, the generality of this analysis makes the semantic value of THAT too unspecific to connect these very general attributive representations to all and only things that are bounded, coherent, and so on.

Burge could try to satisfy condition 2 by holding that THAT stands for an object. This proposal is apparent in Burge's analysis of object segmentation and tracking (2009, sec. 3; 2010, 151, 234). Thus, he could say that object representations are connected to objects because THAT stands for an object. Unfortunately, this approach would prevent him from satisfying condition 1: accommodate the intuition that any constituent of a complex representation, *R*, should make a nonredundant contribution to the content of *R*. After all, if the attributives also refer to objects, what is the point of having a separate constituent that refers to objects too?

In my view, these problems are generated by the misleading linguistic analogy that drives Burge's analysis. Linguistic representations have all-purpose demonstrative representations like 'this' and 'that' that can refer to almost anything. That is why speakers often provide guidance to their interlocutor by pairing these all-purpose devices with attributives; attributives facilitate the task of identifying the referent when there are many equally salient entities to which the demonstrative could refer. As far as I know, there is no evidence that the visual system has all-purpose demonstrative representations that can be made more specific by being paired with general attributives. There is rather evidence for the existence of neurons that covary with specific types of properties and some object constraints that govern visual segmentation and tracking.

It is worth noticing that the problems faced by Burge's account do not even arise when we set aside the linguistic analogy and think of object representations as icons. Indeed, ICA provides a perspicuous characterization of the respective semantic contributions of thin indexes and object constraints. Thin indexes refer to pre-objective elements, and object constraints describe the dispositions of the system to combine active thin indexes in ways that match the ways pre-objective elements are combined in the world. Thus, ICA accommodates the intuition that any constituent of a complex representation, *R*, should make a nonredundant contribution to the content of *R* (condition 1). These dispositions are triggered by the detection of local configurations that are reliably instantiated by objects. Thus, ICA explains why object representations apply to all and only things that are bounded, coherent, and so on (condition 2).

7. Explanatory Advantages. My final argument in favor of ICA is that it has some explanatory advantages over AA. We can appreciate these advantages in light of two problems faced by AA.

Problem 1: Burge (2010, 455 n. 38) rightly points out that a causal relation to a property need not lead to the formation of an object representation. After all, we can stand in causal relations to things other than visual objects: an explosion, a flash of light, or a pile of sand. Thus, something is clearly missing from AA's mechanistic account of the connection between object files and objects. If the mechanism that connects object files with objects is as simple as a brute causal relation, it is unclear why the resulting object file becomes active in response to all and only things that satisfy object constraints. In other words, object constraints do not play any role within AA's account of visual reference.[13]

This problem does not even arise for ICA. On this view, thin indexes single out pre-objective elements that constitute the input to build up object files. Object files are formed when some dispositions are triggered by local configurations of those pre-objective elements. Thus, the reason why object files are not formed in response to paradigmatic explosions, flashes of light, or piles of sand is that the latter are aggregates of pre-objective elements that do not exemplify the local configurations that trigger object constraints.

Problem 2: AA implies that the properties that lead to the formation of object files correspond to "a disjunction of very many properties with nothing in common other than that they attract indexes (i.e., they need not have a definition independent of the FINST mechanism)" (Pylyshyn 2007, 90 n. 12). Hence, the properties we need to explain the activation of object files include 'whatever attracts our attention'. Alas, 'whatever attracts our attention' does not correspond to any natural kind (96). But this is unsatisfactory from a mechanistic perspective. If we want to specify the inputs to a mechanism that activates object files, it would be better to have a well-defined set of properties that can set that mechanism into operation.

Although the properties that cause the activation of thin indexes are disjunctive, they are not as disjunctive as the property denoted by 'whatever attracts our attention'. Suppose that a class of thin indexes becomes active in response to green dots and that some T-junctions trigger the disposition to treat that set of dots as one object. Admittedly, our color receptors fire in response to a disjunction of reflectance properties (Matthen 2005), and there can be many different arrangements that count as T-junctions. Still, these disjunctions of properties are not as heterogeneous as the disjunction of

13. I do not mean to imply that object constraints cannot apply to flashes of light, explosions, or piles of sand. My point is that a purely causal model of visual reference leaves no room for the operation of object constraints. Thus, there is nothing to prevent it from leading to the activation of object files in response to entities that do not satisfy those constraints.

properties formed from 'whatever attracts our attention'. This represents an important step forward in a solution to the grounding problem.

Although the grounding problem is relatively independent from the way philosophers have investigated the naturalization of intentionality, it does have implications for the latter project. Suppose that we think that intentionality should be based on information relations between entities in the world and a representational system. If ICA is correct, these information relations should be introduced to explain the activation of the most basic thin indexes (i.e., those feature detectors that refer in a purely causal manner). One could therefore say that these thin indexes are devices that have the function of carrying information of properties. Still, this would leave the naturalization program seriously incomplete. If we are interested in object representations, we ought to deliver a naturalistic account of the combinations of represented properties that constitute object representations. As a result, we ought to account for the relevant object constraints in naturalistic terms. If these constraints characterize the dispositions of the system to combine basic thin indexes in some ways, we have to provide a naturalistic account of the categorical bases of those dispositions and tell a story about the way our ancestors acquired these dispositions. If it turns out that we cannot provide a nonintentional characterization of those dispositions and their acquisition, this would constitute an important obstacle for the project of incorporating intentionality into a naturalistic worldview.

8. Conclusion. I presented an account of visual reference that conceives of object files as icons. This account enabled me to characterize a family of perceptual errors that are different in kind from the misattribution of properties to objects. These errors arise in segmentation and tracking. I defended this account by showing that it offers a plausible explanation of the empirical evidence that motivates the nonrepresentational account, it can respond to the regress problem, it can offer a plausible solution to the grounding problem, and it has two explanatory advantages over the nonrepresentational account. I also underlined four broader philosophical issues: this debate is ultimately a debate on the explanatory scope of cognitive science, our response to the regress argument can be generalized to recent discussions on the relational character of perceptual experiences, the proposed account constitutes an alternative to Burge's claim that perceptual contents are structured as complex demonstratives, and it introduces a new research avenue in the program of naturalizing intentionality.

REFERENCES

Burge, Tyler. 2009. "Five Theses on *De Re* States and Attitudes." In *The Philosophy of David Kaplan*, ed. Joseph Almog and Paolo Leonardi, 246–324. New York: Oxford University Press.

———. 2010. *Origins of Objectivity*. Oxford: Clarendon.
Campbell, John. 2009. "Consciousness and Reference." In *The Oxford Handbook of Philosophy of Mind*, ed. Brian P. McLaughlin, Ansgar Beckermann, and Sven Walter, 648–62. Oxford: Clarendon.
———. 2012. "Perceiving the Intended Model." In *Perception, Realism, and the Problem of Reference*, ed. Athanassios Raftopoulos and Peter Machamer, 96–122. Cambridge: Cambridge University Press.
Carey, Susan. 2009. *The Origin of Concepts*. New York: Oxford University Press.
Carey, Susan, and Fei Xu. 2001. "Infants' Knowledge of Objects: Beyond Object Files and Object Tracking." *Cognition* 80:179–213.
Dickie, Imogen. 2010. "We Are Acquainted with Ordinary Things." In *New Essays on Singular Thought*, ed. Robin Jeshion, 213–45. Oxford: Clarendon.
Echeverri, Santiago. 2016a. "Illusions of Optimal Motion, Relationism, and Perceptual Content." *Pacific Philosophical Quarterly*, forthcoming. doi:10.1111/papq.12159.
———. 2016b. "Object Files, Properties, and Perceptual Content." *Review of Philosophy and Psychology* 7:283–307.
Fodor, Jerry A. 2008. *LOT 2: The Language of Thought Revisited*. New York: Oxford University Press.
Hoffman, Donald D. 1998. *Visual Intelligence: How We Create What We See*. New York: Norton.
Humphreys, Glyn, and Jane Riddoch. 2014. *A Case Study in Visual Agnosia Revisited: To See but Not to See*. London: Psychology Press.
Kahneman, Daniel, Anne Treisman, and B. Gibbs. 1992. "The Reviewing of Object Files: Object-Specific Integration of Information." *Cognitive Psychology* 24:175–219.
Leslie, Alan M., Fei Xu, P. D. Tremoulet, and Brian J. Scholl. 1998. "Indexing and the Object Concept: Developing 'What' and 'Where' Systems." *Trends in Cognitive Sciences* 2 (1): 10–28.
Matthen, Mohan. 2005. *Seeing, Doing, and Knowing: A Philosophical Theory of Sense Perception*. Oxford: Clarendon.
———. 2014. "Image Content." In *Does Perception Have Content?* ed. Berit Brogaard, 265–90. New York: Oxford University Press.
Millikan, Ruth G. 2000. *On Clear and Confused Ideas*. Cambridge: Cambridge University Press.
Piccinini, Gualtiero. 2015. "Computation in Physical Systems." In *Stanford Encyclopedia of Philosophy*, ed. Eward N. Zalta. Stanford, CA: Stanford University. http://plato.stanford.edu/archives/sum2015/entries/computation-physicalsystems/.
Prinz, Jesse J. 2012. *The Conscious Brain: How Attention Engenders Experience*. New York: Oxford University Press.
Pylyshyn, Zenon W. 1984. *Computation and Cognition: Toward a Foundation of Cognitive Science*. Cambridge, MA: MIT Press.
———. 2001. "Visual Indexes, Preconceptual Objects, and Situated Vision." *Cognition* 80 (1/2): 127–58.
———. 2003. *Seeing and Visualizing: It's Not What You Think*. Cambridge, MA: MIT Press.
———. 2004. "Some Puzzling Findings in Multiple Object Tracking (MOT): Tracking without Keeping Track of Object Identities." *Visual Cognition* 11 (7): 801–22.
———. 2007. *Things and Places: How the Mind Connects with the World*. Cambridge, MA: MIT Press.
Recanati, François. 2012. *Mental Files*. New York: Oxford University Press.
Rensnik, Ronald A. 2000. "The Dynamic Representation of Scenes." *Visual Cognition* 7 (1–3): 17–42.
Scholl, Brian J. 2007. "Object Persistence in Philosophy and Psychology." *Mind and Language* 22 (5): 563–91.
Scholl, Brian J., and Zenon W. Pylyshyn. 1999. "Tracking Multiple Items through Occlusion: Clues to Visual Objecthood." *Cognitive Psychology* 38:259–90.
Sears, Christopher R., and Zenon W. Pylyshyn. 2000. "Multiple Object Tracking and Attentional Processes." *Canadian Journal of Experimental Psychology* 54 (1): 1–14.
Spelke, Elizabeth S. 1988. "Where Perceiving Ends and Thinking Begins: The Apprehension of Objects in Infancy." In *Perceptual Development in Infancy: Minnesota Symposium on Child Psychology*, vol. 20, ed. Albert Yonas, 197–234. Hillsdale, NJ: Erlbaum.
———. 1994. "Initial Knowledge: Six Suggestions." *Cognition* 50:431–45.
Tremoulet, P. D., Alan M. Leslie, and D. Geoffrey Hall. 2000. "Infant Individuation and Identification of Objects." *Cognitive Development* 15 (4): 499–522.